Date: May 5, 1997

EXPRESS MAIL LABEL NO. EM080924599US

Inventors: Aldis Darzins and Gregory T. Mrachko
Attorney's Docket No.: EBC97-06A

A *SPHINGOMONAS* BIODESULFURIZATION CATALYST

RELATED APPLICATIONS

This is a continuation-in-part application filed under 37 CFR 1.53(b)(1) and claims priority to Serial No. _____ filed April 7, 1997, entitled A *SPHINGOMONAS* BIODESULFURIZATION CATALYST by Aldis Darzins and Gregory T. Mrachko, the contents of which are incorporated herein by reference in its entirety.

BACKGROUND OF THE INVENTION

The microbial desulfurization of fossil fuels has been an area of active investigation for over fifty years. The object of these investigations has been to develop biotechnology based methods for the pre-combustion removal of sulfur from fossil fuels, such as coal, crude oil and petroleum distillates. The driving forces for the development of desulfurization methods are the increasing levels of sulfur in fossil fuel and the increasingly stringent regulation of sulfur emissions. Monticello *et al.*, "Practical Considerations in Biodesulfurization of

Petroleum," IGT's 3d Intl. Symp. on Gas, Oil, Coal and Env. Biotech., (Dec. 3-5, 1990) New Orleans, LA.

Many biocatalysts and processes have been developed to desulfurize fossil fuels, including those described in U.S. Patent Nos. 5,356,801, 5,358,870, ~~5,358,813~~ 5,356,813, 5,198,341, 5,132,219, 5,344,778, 5,104,801 and 5,002,888, incorporated herein by reference. Economic analyses indicate that one limitation in the commercialization of the technology is improving the reaction rates and specific activities of the biocatalysts, such as the bacteria and enzymes that are involved in the desulfurization reactions. The reaction rates and specific activities (sulfur removed/hour/gram of biocatalyst) that have been reported in the literature are much lower than those necessary for optimal commercial technology. Therefore, improvements in the longevity and specific activity of the biocatalyst are desirable.

SUMMARY OF THE INVENTION

The invention relates to a novel microorganism, designated *Sphingomonas* sp. strain AD109, as well as isolated proteins and nucleic acid sequences obtained from this microorganism. This microorganism was obtained using a soil enrichment process using 2-(2-hydroxyphenyl)benzenesulfinate as the sole sulfur source. A biologically pure sample of this microorganism has been isolated and characterized.

The invention also relates to a collection of desulfurization enzymes isolated from *Sphingomonas* sp. strain AD109 which, together, catalyze the oxidative desulfurization of dibenzothiophene (DBT).

In another embodiment, the invention includes an isolated nucleic acid molecule, such as a DNA or RNA nucleotide sequence or molecule, which encodes one or more of the *Sphingomonas* desulfurization enzymes, or a homologue

or active fragment thereof. The invention also includes a recombinant microorganism containing one or more heterologous nucleic acid molecules which encode one or more of the *Sphingomonas* desulfurization enzymes or homologues or active fragments thereof.

In a further embodiment, the invention provides a method of using the *Sphingomonas* microorganism or an enzyme preparation derived therefrom as a biocatalyst in the biocatalytic desulfurization of a fossil fuel containing organosulfur compounds. The method comprises the steps of (1) contacting the fossil fuel with an aqueous phase containing a *Sphingomonas* biocatalyst which is capable of biocatalytic desulfurization and, optionally, a flavoprotein, thereby forming a fossil fuel and aqueous phase mixture; (2) maintaining the mixture under conditions sufficient for sulfur oxidation and/or cleavage of the carbon-sulfur bonds of the organosulfur molecules by the biocatalyst, and (3) separating the fossil fuel having a reduced organic sulfur content from the resulting aqueous phase.

The invention also provides a method of oxidizing an organic compound. The method comprises the steps of: (1) contacting the organic compound with an aqueous phase containing a *Sphingomonas* biocatalyst comprising at least one enzyme capable of catalyzing at least one step in the oxidative cleavage of carbon-sulfur bonds, thereby forming an organic compound and aqueous phase mixture; (2) maintaining the mixture of step (1) under conditions sufficient for oxidation of the organic compound by the biocatalyst, thereby resulting in an oxidized organic compound, and, optionally, separating the oxidized organic compound from the aqueous phase.

4

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A, 1B, 1C and 1D together set forth the DNA sequence and the corresponding amino acid sequence of open reading frame 1 (ORF-1, *dszA*) of the nucleotide sequence required for desulfurization activity in *Sphingomonas* sp. strain AD109.

Figures 2A, 2B and 2C together set forth the DNA sequence and the corresponding amino acid sequence of open reading frame 2 (ORF-2, *dszB*) of the nucleotide sequence required for desulfurization activity in *Sphingomonas* sp. strain AD109.

Figures 3A, 3B and 3C together set forth the DNA sequence and the corresponding amino acid sequence of open reading frame 3 (ORF-3, *dszC*) of the nucleotide sequence required for desulfurization activity in *Sphingomonas* sp. strain AD109.

Figure 4 is a graph showing the disappearance of 2-(2-phenyl)benzenesulfinate (HPBS) and the appearance of 2-hydroxybiphenyl (2-HBP) in the presence of *Sphingomonas* AD109 cell-free lysates.

Figure 5 shows a physical map of the *Sphingomonas dsz* gene cluster.

Figures 6A, 6B, 6C, 6D, 6E, 6F and 6G together set forth the nucleotide sequence of the *Sphingomonas dsz* gene cluster.

Figure 7 is a physical map of the plasmid pDA296.

Figure 8 presents the results of a GAP analysis of the DszA proteins from *Sphingomonas* sp. strain AD109 and *Rhodococcus* IGTS8.

Figure 9 presents the results of a GAP analysis of the DszB proteins from *Sphingomonas* sp. strain AD109 and *Rhodococcus* IGTS8.

Figure 10 presents the results of a GAP analysis of the sequences of the DszC proteins from *Sphingomonas* sp. strain AD109 and *Rhodococcus* IGTS8.

Figure 11 is a physical map of the plasmid pEBCtac.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is based on the discovery and isolation of a novel microorganism which is capable of selectively desulfurizing dibenzothiophene ("DBT"). As described in Example 1, this microorganism was obtained from soil samples obtained at sites contaminated with petroleum and petroleum by-products by a soil enrichment procedure using 2-(2-hydroxyphenyl)benzenesulfinate as the sole sulfur source. A biologically pure sample of the novel microorganism has been isolated and characterized. The microorganism is a motile, gram-negative rod. Based on a fatty acid analysis, as described in Example 2, this microorganism has been identified as a *Sphingomonas* species, and designated strain AD-109. This microorganism has been deposited at the American Type Culture Collection (ATCC), 12301 Park Lawn Drive, Rockville, Maryland, U.S.A. 20852 under the terms of the Budapest Treaty and has been designated as ATCC Deposit No. 55954 on April 21, 1997.

The novel microorganism of the invention can be grown by fermentation under aerobic conditions in the presence of a sulfur-free mineral salts medium (e.g., 4 g/L K_2HPO_4, 4 g/L Na_2HPO_4, 2 g/L NH_4Cl, 0.2 g/L $MgCl_2.6H_2O$, 0.001 g/L $CaCl_2.2H_2O$, and 0.001 g/L $FeCl_3.6H_2O$), containing a sulfur-free source of assimilable carbon such as glucose. The sole source of sulfur provided can be a heterocyclic organosulfur compound, such as dibenzothiophene or a derivative thereof.

Sphingomonas sp. strain AD109 expresses a collection of enzymes which together catalyze the conversion of DBT to

2-hydroxybiphenyl (also referred to as "2-HBP") and inorganic sulfur. An enzyme which catalyzes one or more steps in this overall process is referred to herein as a "desulfurization enzyme". The nucleic acid sequence required for this overall process has been identified and cloned using the general method described in U.S. Patent No. 5,356,801, the contents of which are incorporated herein by reference, and is set forth in Figure 6 (SEQ ID NO.: 12). This nucleic acid sequence (also referred to as the "*Sphingomonas* dsz sequence") comprises three open reading frames, designated ORF-1 (base pairs 442-1800, also set forth in Figures 1A-1D and SEQ ID NO.: 1), ORF-2 (base pairs 1800-2909, also set forth in Figures 2A-2C and SEQ ID NO.: 3) and ORF-3 (base pairs 2906-4141, sequence also set forth in Figures 3A-3C and SEQ ID NO.: 5). The predicted amino acid sequences encoded by these open reading frames are set forth in Figures 1A-1D (ORF-1, SEQ ID NO: 2), Figures 2A-2C (ORF-2, SEQ ID NO.: 4) and Figures 3A-3C (ORF-3, SEQ ID NO.: 6). Each of these open reading frames is homologous to the corresponding open reading frame of *Rhodococcus* sp. IGTS8; the sequences of the *Rhodococcus* open reading frames are disclosed in U.S. Patent No. 5,356,801.

In one embodiment, the present invention provides an isolated nucleic acid molecule comprising one or more nucleotide sequences which encode one or more of the biodesulfurization enzymes of *Sphingomonas* sp. strain AD109. The isolated nucleic acid molecule can be, for example, a nucleotide sequence, such as a deoxyribonucleic acid (DNA) sequence or a ribonucleic acid (RNA) sequence. Such a nucleic acid molecule comprises one or more nucleotide sequences which encode one or more of the amino acid sequences set forth in SEQ ID NO.: 2, SEQ ID NO.: 4, and SEQ ID NO.: 6. For example, the isolated nucleic acid

molecule can comprise one or more of the nucleotide sequences of SEQ ID NO.: 1, SEQ ID NO.: 3, and SEQ ID NO.: 5, or a complement of any of these sequences. The isolated nucleic acid molecule can also comprise a nucleotide sequence which results from a silent mutation of one or more of the sequences set forth in SEQ ID NO.: 1, SEQ ID NO.: 3, and SEQ ID NO.: 5. Such a nucleotide sequence can result, for example, from a mutation of the native sequence in which one or more codons have been replaced with a degenerate codon, i.e., a codon which encodes the same amino acid. Such mutant nucleotide sequences can be constructed using methods which are well known in the art, for example the methods discussed by Ausubel *et al.*, *Current Protocols in Molecular Biology*, Wiley-Interscience, New York (1997)(hereinafter "Ausubel *et al.*") and by Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, third edition, Cold Spring Harbor Laboratory Press (1992) (hereinafter "Sambrook *et al.*"), each of which are incorporated herein by reference.

In another embodiment, the invention includes an isolated nucleic acid molecule comprising a nucleotide sequence which is homologous to one or more of the sequences of SEQ ID NO.: 1, SEQ ID NO.: 3, and SEQ ID NO.: 5, or complements thereof. Such a nucleotide sequence exhibits at least about 80% homology, or sequence identity, with one of these *Sphingomonas* nucleotide sequences, preferably at least about 90% homology or sequence identity. Particularly preferred sequences have at least about 95% homology or have essentially the same sequence. Preparation of mutant nucleotide sequences can be accomplished by methods known in the art as are described in Old, *et al.*, *Principles of Gene Manipulation*, Fourth

Edition, Blackwell Scientific Publications (1989), in Sambrook *et al.*, and in Ausubel *et al.*

The invention further includes nucleic acid molecules which are useful as hybridization probes, for example, for the isolation of the *Sphingomonas* genes encoding desulfurization enzymes or identical or homologous genes from other organisms. Such molecules comprise nucleotide sequences which hybridize to all or a portion of the nucleotide sequence of SEQ ID NO.: 1, SEQ ID NO.: 3 or SEQ ID NO.: 5 or to non-coding regions immediately (within about 1000 nucleotides) 5' or 3' of each open reading frame. The invention also includes an isolated nucleic acid molecule which comprises a fragment of one or more of the nucleotide sequences set forth in SEQ ID NO.: 1, SEQ ID NO.: 3 or SEQ ID NO.: 5 or complements of any of these sequences. Such a fragment will generally comprise at least about 20 or at least about 40 contiguous nucleotides and, preferably, at least about 50 contiguous nucleotides of one of the disclosed sequences. Preferably, the hybridization probe of the invention hybridizes to one of these sequences under stringent conditions, such as those set forth by Sambrook *et al*. and Ausubel *et al*. For example, under conditions of high stringency, such as high temperatures and low salt concentrations, only DNA molecules which are essentially exact matches, or complements, will hybridize, particularly if the probe is relatively short. Hybridization under conditions of lower stringency, such as low temperatures, low formamide concentrations and high salt concentrations, allows greater mismatch between the probe and the target DNA molecule. It is particularly preferred that the nucleic acid molecule hybridizes selectively to the disclosed sequence(s).

The nucleic acid molecules can be synthesized chemically from the disclosed sequences. Alternatively,

the nucleic acid molecules can be isolated from a suitable nucleic acid library (such as a DNA library) obtained from a microorganism which is believed to possess the nucleic acid molecule (such as, *Sphingomonas* sp. strain AD109), employing hybridizing primers and/or probes designed from the disclosed sequences. Such a method can result in isolating the disclosed molecules (or spontaneous mutants thereof) for use in preparing recombinant enzymes, confirming the disclosed sequences, or for use in mutagenizing the native sequences.

In yet another embodiment, the nucleic acid molecule of the present invention can be a nucleic acid molecule, such as a recombinant DNA molecule, resulting from the insertion into its chain by chemical or biological means, of one or more of the nucleotide sequences described above. Recombinant DNA includes any DNA synthesized by procedures using restriction nucleases, nucleic acid hybridization, DNA cloning, DNA synthesis or any combination of the preceding. Methods of construction can be found in Sambrook *et al.* and Ausubel *et al.*, and additional methods are known by those skilled in the art.

The isolated nucleic acid molecule of the invention can further comprise a nucleotide sequence which encodes an oxidoreductase, such as a flavoprotein, such as a flavin reductase. For example, the nucleic acid molecule can encode an oxidoreductase which is native to *Sphingomonas* sp. strain AD109. The nucleic acid molecule can also encode the oxidoreductase denoted DszD described in copending U.S. Patent Application Serial No. 08/583,118; the flavin reductase from *Vibrio harveyii* described in copending U.S. Patent Application Serial No. 08/351,754; or the flavin reductase from *Rhodococcus* sp. IGTS8, described in copending U.S. Patent Application Serial No. 08/735,963.

The contents of each of these applications are incorporated herein by reference.

The invention also includes a plasmid or vector comprising a recombinant DNA sequence or molecule which comprises one or more of the nucleic acid molecules, e.g. nucleotide sequences, of the invention, as described above. The terms "plasmid" and "vector" are intended to encompass any replication competent or replication incompetent plasmid or vector capable of having foreign or exogenous DNA inserted into it by chemical or biological means and subsequently, when transformed into an appropriate non-human host organism, of expressing the product of the foreign or exogenous DNA insert (e.g., of expressing the biocatalyst and flavoprotein of the present invention). In addition, the plasmid or vector is receptive to the insertion of a DNA molecule or fragment thereof containing the gene or genes of the present invention, said gene or genes encoding a biocatalyst as described herein. Procedures for the construction of DNA plasmid vectors include those described in Sambrook *et al.* and Ausubel *et al.* and others known by those skilled in the art.

The plasmids of the present invention include any DNA fragment containing a nucleotide sequence as described above. The DNA fragment should be transmittable, for example, to a host microorganism by transformation or conjugation. Procedures for the construction or extraction of DNA plasmids include those described in Sambrook *et al.* and Ausubel *et al.*, and others known by those skilled in the art. In one embodiment, the plasmid comprises a nucleotide sequence of the invention operatively linked to a competent or functional regulatory sequence. Examples of suitable regulatory sequences include promoters, enhancers, transcription binding sites, ribosomal binding sites, transcription termination sequences, etc.

In one preferred embodiment, the regulatory or promoter sequences are those native to the *Sphingomonas* operon containing the genes disclosed herein. In yet another embodiment, one or more regulatory sequences (e.g. the promoter) is native to the selected host cell for expression. The promoter can be selected so that the gene or genes are inducible or constitutively expressed. Furthermore, the sequences can be regulated individually or together, as an operon. Examples of suitable promoters include the *E. coli lac* and *tac* promoters and the Pseudomonas P_G promoter (Yen, *J. Bacteriol.* **173** : 5328-5335 (1991)). An example of such a plasmid and its construction are described in Example 8.

In another embodiment, the invention relates to a recombinant or transformed non-human host organism which contains a heterologous DNA molecule of the invention as described above. The recombinant non-human host organisms of the present invention can be created by various methods by those skilled in the art. Any method for introducing a recombinant plasmid, such as a plasmid of the invention described above, into the organism of choice can be used, and a variety of such methods are described by Sambrook *et al.* and Ausubel *et al.* For example, the recombinant plasmid can be introduced via a suitable vector by transformation, conjugation, transduction or electroporation. By the term "non-human host organism" is intended any non-human organism capable of the uptake and expression of foreign, exogenous or recombinant DNA.

The recombinant microorganism can be derived from a host organism which does not express a native desulfurization biocatalyst. Such microorganisms include bacteria and yeasts, e.g., *E. coli, Bacillus,* and non-desulfurizing pseudomonads (as described in U. S. Patent

aa Application Serial Number 08/851,085 ~~(Attorney Docket No. EBC96-06A, being filed concurrently herewith))~~. In another embodiment, the recombinant microorganism is derived from a host organism which expresses a native biodesulfurization catalyst. Preferred microorganisms of this type are *Rhodococcus* sp. IGTS8 (ATCC 53968), recombinant microorganisms comprising one or more of the IGTS8 desulfurizing genes and *Sphingomonas* sp. strain AD109. Other desulfurizing microorganisms which are suitable host organisms include *Corynebacterium* sp. strain SY1, as disclosed by Omori *et al.*, *Appl. Env. Microbiol.*, **58** : 911-915 (1992); *Rhodococcus erythropolis* D-1, as disclosed by Izumi *et al.*, *Appl. Env. Microbiol.*, **60** :223-226 (1994); the *Arthrobacter* strain described by Lee *et al.*, *Appl. Environ. Microbiol.* **61** : 4362-4366 (1995); the *Agrobacterium* strain disclosed by Constanti *et al.*, *Enzyme Microb. Tech.* **19** : 214-219 (1996) and the *Rhodococcus* strains (ATCC 55309 and ATCC 55310) disclosed by Grossman *et al.*, U.S. Patent No. 5,607,857, each of which is incorporated herein by reference in its entirety. Each of these microorganisms produces one or more enzymes (protein biocatalysts) that catalyze one or more reactions in the desulfurization of DBT.

The invention also relates to desulfurization enzymes which can be isolated from *Sphingomonas* sp. strain AD109. These include desulfurization enzymes which catalyze one or more steps in the oxidative desulfurization of DBT. The enzyme encoded by ORF-2 has been partially purified and exhibits 2-(2-hydroxyphenyl)benzenesulfinate (HPBS) desulfinase activity and has an apparent molecular weight by denaturing gel electrophoresis of about 40,000 daltons.

In one embodiment, the invention includes an isolated desulfurization enzyme from *Sphingomonas* sp. strain AD109 using methods and assays which are known the art, for example, the methods used by Gray *et al.* to isolate and characterize desulfurization enzymes from *Rhodococcus* IGTS8 (Gray *et al.*, *Nature Biotech.* **14** : 1705-1709 (1996)). These enzymes can be isolated or purified from the cell by lysing the cell and subjecting the cell lysate to known protein purification methods, and testing the fractions obtained thereby for the desired enzymatic activity. Examples of suitable protein purification methods include ammonium sulfate precipitation, ultrafiltration, diafiltration, immunoabsorption, anion exchange chromatography, gel filtration chromatography and hydrophobic interaction chromatography. The enzymes of the invention can also be recombinant proteins produced by heterologous expression of a nucleic acid molecule comprising a nucleotide sequence of SEQ ID NO.: 1, SEQ ID NO.: 3 or SEQ ID NO.: 5; or a mutation or fragment thereof, as discussed above. When the recombinant organism is derived from a non-*Sphingomonas* host, the recombinant proteins can be prepared in a form which is substantially free of other *Sphingomonas* proteins.

The invention also includes an isolated enzyme having an amino acid sequence which is homologous to the amino acid sequence of SEQ ID NO.: 2, SEQ ID NO.: 4 or SEQ ID NO.: 6, or fragments thereof. The term "homologous" or "homologue", as used herein, describes a protein (which is not obtained from *Rhodococcus* or *Rhodococcus* sp IGTS8) having at least about 80% sequence identity or homology with the reference protein, and preferably about 90% sequence homology, in an amino acid alignment. Most preferably, the protein exhibits at least about 95%

homology or essentially the same sequence as the disclosed sequence. An amino acid alignment of two or more proteins can be produced by methods known in the art, for example, using a suitable computer program, such as BLAST (Altschul *et al.*, *J. Mol. Biol.* **215** : 403-410 (1990)). A homologous protein can also have one or more additional amino acids appended at the carboxyl terminus or amino terminus, such as a fusion protein.

The homologous enzymes described herein can be native to an organism, such as a desulfurizing microorganism, including *Sphingomonas* sp. strain AD109 and mutants thereof. Such enzymes can be isolated from such sources using standard techniques and assays, as are described in the Exemplification and others known in the art. For example, the *Sphingomonas* desulfurization enzymes can be used to induce the formation of antibodies, such as monoclonal antibodies, according to known methods. The antibodies can then be used to purify the desulfurization enzymes from a desulfurizing organism via affinity chromatography, as is well known in the art.

The homologous enzymes of the invention can also be non-naturally occurring. For example, a homologous enzyme can be a mutant desulfurization enzyme which has a modified amino acid sequence resulting from the deletion, insertion or substitution of one or more amino acid residues in the amino acid sequence of a *Sphingomonas* desulfurization enzyme. Such amino acid sequence variants can be prepared by methods known in the art. For example, the desired polypeptide can be synthesized *in vitro* using known methods of peptide synthesis. The amino acid sequence variants are preferably made by introducing appropriate nucleotide changes into a DNA molecule encoding the native enzyme, followed by expression of the mutant enzyme in an

appropriate vector, such as *E. coli*. These methods include site-directed mutagenesis or random mutagenesis, for example.

Particularly preferred mutants include those having amino acid sequences which include the amino acid residues which are encoded by both SEQ ID NO.: 1, SEQ ID NO.: 3 or SEQ ID NO.: 5 and the corresponding open reading frame of *Rhodococcus* sp. IGTS8, as disclosed in U.S. Patent No. 5,356,801. That is, these mutants include the amino acid residues which are conserved in these two organisms in an amino acid alignment. Mutants which result from conservative substitution of one or more of these conserved residues, as well as non-conserved residues, are also included. Conservative and non-conservative substitutions (including deletions and insertions) can be made in non-conserved regions of the amino acid sequence and mutants resulting from both conservative and non-conservative substitutions of these residues are included herein.

Conservative substitutions are those in which a first amino acid residue is substituted by a second residue having similar side chain properties. An example of such a conservative substitution is replacement of one hydrophobic residue, such as valine, with another hydrophobic residue, such as leucine. A non-conservative substitution involves replacing a first residue with a second residue having different side chain properties. An example of this type of substitution is the replacement of a hydrophobic residue, such as valine, with an acidic residue, such as glutamic acid.

The two primary variables in the construction of amino acid sequence variants are (1) the location of the mutation site and (2) the nature of the mutation. These variables can be manipulated to identify amino acid residues at the active site of the enzyme. For example, an amino acid

substitution which yields a mutant enzyme having significantly different activity than the native enzyme suggests that the substituted amino acid residue is at the active site. Such mutants can have the same or similar, increased or decreased activity relative to that of the native enzyme.

Amino acids can be modified, for example, by substituting first with a conservative choice, followed by non-conservative choices depending upon the results achieved, by deleting the target residue(s) or by inserting residues adjacent to a particular site. Variants can also be constructed using a combination of these approaches.

The proteins of the present invention can be produced using techniques to overexpress the gene, as are described by Sambrook *et al.* and Ausubel *et al.* Improved expression, activity or overexpression of the *Sphingomonas* desulfurization enzymes (in *Sphingomonas* sp AD 109 or in recombinant host cells harboring the disclosed nucleic acid molecules) can also be accomplished by mutagenesis. Suitable mutagens include radiation, e.g., ultraviolet radiation, and chemical mutagens, such as N-methyl-N'-nitroso-guanidine, hydroxylamine, ethylmethanesulfonate and nitrous acid. Furthermore, spontaneous mutants can be selected where the microorganism is subjected to an enrichment culture, as exemplified herein. The mutagenesis and subsequent screening for mutants harboring increased enzymatic activity can be conducted according to methods generally known in the art.

The present invention also provides a method of desulfurizing a carbonaceous material containing organosulfur molecules. The carbonaceous material can be, for example, a DBT-containing material or a fossil fuel, such as petroleum, a petroleum distillate fraction or coal. The method comprises the steps of (1) contacting the

carbonaceous material with an aqueous phase containing a *Sphingomonas*-derived biocatalyst comprising at least one enzyme capable of catalyzing at least one step in the oxidative cleavage of carbon-sulfur bonds, thereby forming a carbonaceous material and aqueous phase mixture; (2) maintaining the mixture of step (1) under conditions sufficient for biocatalysis; and (3) separating the carbonaceous material having a reduced organic sulfur content from the resulting aqueous phase.

The term "*Sphingomonas*-derived biocatalyst", as used herein, is a biocatalyst which includes one or more desulfurization enzymes encoded by SEQ ID NO.: 1, SEQ ID NO.: 3 and SEQ ID NO.: 5; or a mutant or homologue thereof. In one embodiment, the biocatalyst is a microorganism, such as *Sphingomonas sp.* strain AD109. The biocatalyst can also be a recombinant organism which contains one or more heterologous nucleotide sequences or nucleic acid molecules as described above.

Although living microorganisms (e.g., a culture) can be used as the biocatalyst herein, this is not required. Biocatalytic enzyme preparations that are useful in the present invention include microbial lysates, extracts, fractions, subfractions, or purified products obtained by conventional means and capable of carrying out the desired biocatalytic function. Generally, such enzyme preparations are substantially free of intact microbial cells. In a particularly preferred embodiment, the biocatalyst is overexpressed in the recombinant host cell (such as a cell which contains more than one copy of the gene or genes).

Enzyme biocatalyst preparations suitable for use herein can optionally be affixed to a solid support, e.g., a membrane, filter, polymeric resin, glass particles or beads, or ceramic particles or beads. The use of immobilized enzyme preparations facilitates the separation

of the biocatalyst from the treated fossil fuel which has been depleted of refractory organosulfur compounds.

A fossil fuel that is suitable for desulfurization treatment according to the present invention is one that contains organic sulfur. Such a fossil fuel is referred to as a "substrate fossil fuel". Substrate fossil fuels that are rich in thiophenic sulfur are particularly suitable for desulfurization according to the method described herein. Examples of such substrate fossil fuels include Cerro Negro or Orinoco heavy crude oils; Athabascan tar and other types of bitumen; petroleum refining fractions such as gasoline, kerosene, diesel, fuel oil, residual oils and miscellaneous refinery by-products; shale oil and shale oil fractions; and coal-derived liquids manufactured from sources such as Pocahontas #3, Lewis-Stock, Australian Glencoe or Wyodak coal.

In the petroleum extraction and refining arts, the term "organic sulfur" is generally understood as referring to organic molecules having a hydrocarbon framework to which one or more sulfur atoms are covalently joined. These sulfur atoms can be directly bonded to the hydrocarbon framework, e.g., by one or more carbon-sulfur bonds, or can be present in a substituent bonded to the hydrocarbon framework of the molecule, e.g., a sulfate group. The general class of organic molecules having one or more sulfur heteroatoms are sometimes referred to as "organosulfur compounds". The hydrocarbon portion of these compounds can be aliphatic and/or aromatic.

Sulfur-bearing heterocycles, such as substituted and unsubstituted thiophene, benzothiophene, and dibenzothiophene, are known to be stable to conventional desulfurization treatments, such as hydrodesulfurization (HDS). Sulfur-bearing heterocycles can have relatively simple or relatively complex chemical structures. In complex heterocycles, multiple condensed aromatic rings,

one or more of which can be heterocyclic, are present. The difficulty of desulfurization generally increases with the structural complexity of the molecule. That is, refractory behavior is particularly accentuated in complex sulfur-bearing heterocycles, such as dibenzothiophene (DBT, $C_{12}H_8S$).

Much of the residual post-HDS organic sulfur in fossil fuel refining intermediates and combustible products is thiophenic sulfur. The majority of this residual thiophenic sulfur is present in DBT and derivatives thereof having one or more alkyl or aryl groups attached to one or more carbon atoms present in one or both flanking benzo rings. DBT itself is accepted as a model compound illustrative of the behavior of the class of compounds encompassing DBT and derivatives thereof in reactions involving thiophenic sulfur (Monticello and Finnerty, *Ann. Rev. Microbiol.*, **39** : 371-389 (1985)). DBT and derivatives thereof can account for a significant percentage of the total sulfur content of particular crude oils, coals and bitumen. For example, these sulfur-bearing heterocycles have been reported to account for as much as 70 wt% of the total sulfur content of West Texas crude oil, and up to 40 wt% of the total sulfur content of some Middle East crude oils. Thus, DBT is considered to be particularly relevant as a model compound for the forms of thiophenic sulfur found in fossil fuels, such as crude oils, coals or bitumen of particular geographic origin, and various refining intermediates and fuel products manufactured therefrom (Monticello and Finnerty (1985), *supra*). Another characteristic of DBT and derivatives thereof is that, following a release of fossil fuel into the environment, these sulfur-bearing heterocycles persist for long periods of time without significant biodegradation. Gundlach *et al.*, *Science* **221** : 122-129 (1983). Thus, most prevalent naturally

occurring microorganisms do not effectively metabolize and break down sulfur-bearing heterocycles.

Biocatalytic desulfurization (biocatalysis or BDS) is the excision (liberation or removal) of sulfur from organosulfur compounds, including refractory organosulfur compounds such as sulfur-bearing heterocycles, as a result of the oxidative, preferably selective, cleavage of carbon-sulfur bonds in said compounds by a biocatalyst. BDS treatment yields the desulfurized combustible hydrocarbon framework of the former refractory organosulfur compound, along with inorganic sulfur substances which can be readily separated from each other by known techniques such as fractional distillation or water extraction. For example, DBT is converted into 2-hydroxybiphenyl when subjected to BDS treatment. A suitable biocatalyst for BDS comprises *Sphingomonas* sp. strain AD109 or an enzyme preparation derived therefrom, optionally, in combination with one or more additional non-human desulfurizing organisms (e.g., microorganisms); or an enzyme preparation derived from such an organism. Suitable additional desulfurizing organisms include those described above.

The specific activity of a given biocatalyst is a measure of its biocatalytic activity per unit mass. Thus, the specific activity of a particular biocatalyst depends on the nature or identity of the microorganism used or used as a source of biocatalytic enzymes, as well as the procedures used for preparing and/or storing the biocatalyst preparation. The concentration of a particular biocatalyst can be adjusted as desired for use in particular circumstances. For example, where a culture of living microorganisms, such as *Sphingomonas* sp. strain AD109, is used as the biocatalyst preparation, a suitable culture medium lacking a sulfur source other than sulfur-bearing heterocycles can be inoculated with suitable microorganisms

and grown until a desired culture density is reached. The resulting culture can be diluted with additional medium or another suitable buffer, or microbial cells present in the culture can be retrieved e.g., by centrifugation, and resuspended at a greater concentration than that of the original culture. The concentrations of microorganism and enzyme biocatalyst can be adjusted similarly. In this manner, appropriate volumes of biocatalyst preparations having predetermined specific activities and/or concentrations can be obtained.

In the biocatalytic desulfurization stage, the liquid fossil fuel containing sulfur-bearing heterocycles is combined with the biocatalyst. The relative amounts of biocatalyst and liquid fossil fuel can be adjusted to suit particular conditions, or to produce a particular level of residual sulfur in the treated, deeply desulfurized fossil fuel. The amount of biocatalyst preparation to be combined with a given quantity of liquid fossil fuel will reflect the nature, concentration and specific activity of the particular biocatalyst used, as well as the nature and relative abundance of inorganic and organic sulfur compounds present in the substrate fossil fuel and the degree of deep desulfurization sought or considered acceptable.

The method of desulfurizing a fossil fuel of the present invention involves two aspects. First, a host organism or biocatalytic preparation obtained therefrom is contacted with a fossil fuel to be desulfurized. This can be done in any appropriate container, optionally fitted with an agitation or mixing device. The mixture is combined thoroughly and maintained or allowed to incubate for a sufficient time to allow for biocatalysis. In one embodiment, an aqueous emulsion or microemulsion is produced with an aqueous culture of the organism or enzyme fraction and the fossil fuel, allowing the organism to

propagate in the emulsion while the expressed biocatalyst cleaves carbon-sulfur bonds.

Variables such as temperature, pH, oxidation levels, mixing rate and rate of desulfurization will vary according to the nature of the biocatalyst used. Optimal parameters can generally be determined through no more than routine experimentation.

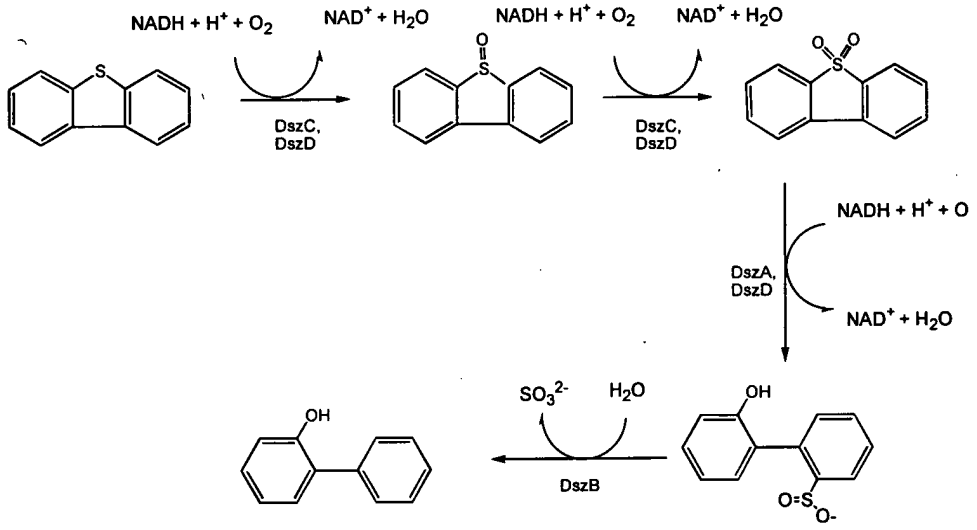
When the fossil fuel is a liquid hydrocarbon, such as petroleum, the desulfurized fossil fuel and the aqueous phase can form an emulsion. The components of such emulsions can be separated by a variety of methods, such as those described in U.S. Patent No. 5,358,870 and U.S. Patent Application Serial No. 08/640,129, which are incorporated herein by reference. For example, some emulsions reverse spontaneously when maintained under stationary conditions for a suitable period of time. Other emulsions can be reversed by adding an additional amount of an aqueous phase. Still other emulsions can be separated by the addition of a suitable chemical agent, such as a demulsifying agent or by employing suitable physical conditions, such as a particular temperature range.

The biocatalyst can be recovered from the aqueous phase, for example, by centrifugation, filtration or lyophilization. When the biocatalyst is a microorganism, the biocatalyst can be resuspended in fresh sulfur-free nutrient medium and/or any fresh microorganism culture as necessary to reconstitute or replenish to the desired level of biocatalytic activity. The biocatalyst can then be reintroduced into the reaction system.

Several suitable techniques for monitoring the rate and extent of desulfurization are well-known and readily available to those skilled in the art. Baseline and time course samples can be collected from the incubation mixture, and prepared for a determination of the residual organic sulfur in the fossil fuel. The disappearance of

sulfur from organosulfur compounds, such as DBT, in the sample being subjected to biocatalytic treatment can be monitored using, e.g., X-ray fluorescence (XRF) or atomic emission spectrometry (flame spectrometry). Preferably, the molecular components of the sample are first separated, e.g., by gas chromatography.

Without being limited to any particular mechanism or theory, it is believed that the pathway of the desulfurization reaction in *Sphingomonas* sp. strain AD109 and other desulfurizing organisms, such as *Rhodococcus* sp. IGTS8, is set forth below:



Here the flavin reductase provides an electron transport chain which delivers, via $FMNH_2$, the reducing equivalents from NADH (or other electron donor) to the enzymes DszC and/or DszA. The enzyme DszC is responsible for the biocatalysis of the oxidation reaction of DBT to $DBTO_2$. The enzyme DszA is responsible for the reaction of $DBTO_2$ to

2-(2-hydroxyphenyl)benzenesulfinate (HPBS). The enzyme DszB catalyzes the conversion of HPBS to 2-hydroxybiphenyl and inorganic sulfur.

Another method of use of the *Sphingomonas* desulfurization enzymes, or mutants, homologues or active fragments thereof, is as a biocatalyst for the oxidation of organic compounds, such as substituted or unsubstituted dibenzothiophenes. The method comprises the steps of (1) contacting the organic compound with an aqueous phase containing a *Sphingomonas*-derived biocatalyst comprising at least one enzyme capable of catalyzing at least one step in the oxidative cleavage of carbon-sulfur bonds, thereby forming an organic compound and aqueous phase mixture; (2) maintaining the mixture of step (1) under conditions sufficient for oxidation of the organic compound by the biocatalyst, thereby resulting in an oxidized organic compound, and, optionally, separating the oxidized organic compound from the aqueous phase. In one embodiment, the organic compound is a heteroorganic compound, such as an organonitrogen compound or an organosulfur compound. In one embodiment, the organic compound is an organosulfur compound which is a component of a fossil fuel, such as petroleum or a petroleum distillate fraction. In a second embodiment, the organic compound is a substituted or unsubstituted indole, as described in U.S. Provisional Patent Application Serial Number 60/020563, filed July 2, 1996, which is incorporated herein by reference.

The enzyme encoded by the nucleotide sequence of ORF-3 catalyzes the oxidation of dibenzothiophene to dibenzothiophene-5,5-dioxide (dibenzothiophene sulfone), and the enzyme encoded by the nucleotide sequence of ORF-1 catalyzes the oxidation of dibenzothiophene-5,5-dioxide to 2-(2-hydroxyphenyl)benzenesulfinate (also referred to as "HPBS"). In one embodiment the biocatalyst comprises the

enzyme encoded by ORF-3, or a mutant, homologue or active fragment thereof; the organosulfur compound is substituted or unsubstituted dibenzothiophene; and the oxidized organosulfur is a substituted or unsubstituted dibenzothiophene-5,5-dioxide or dibenzothiophene-5-oxide (dibenzothiophene sulfoxide). In another embodiment the biocatalyst comprises the enzymes encoded by ORF-1 and ORF-3, or a mutant, homologue or active fragment thereof; the organosulfur compound is a substituted or unsubstituted dibenzothiophene; and the oxidized organosulfur compound is a substituted or unsubstituted 2-(2-hydroxyphenyl)benzenesulfinate. In yet another embodiment, the biocatalyst comprises the enzyme encoded by ORF-1 or a mutant, homologue or active fragment thereof; the organosulfur compound is a substituted or unsubstituted dibenzothiophene-5,5-dioxide; and the oxidized organosulfur compound is a substituted or unsubstituted 2-(2-hydroxyphenyl)benzenesulfinate.

The oxidized organosulfur compound can, optionally, be further processed, for example, via a non-biological process or an enzyme-catalyzed reaction. In one embodiment, the oxidized organosulfur compound is desulfurized in a process employing suitable desulfurization enzymes from an organism other than a *Sphingomonas*.

The biocatalyst can be an organism, such as *Sphingomonas* sp. strain AD109, a desulfurizing mutant thereof, or a recombinant organism or enzyme preparation, as discussed above. When the organosulfur compound is a component of a fossil fuel, suitable reaction conditions and fossil fuel sources can be determined as described above.

The invention will now be further illustrated by the way of the following examples.

EXAMPLES

General Methods and Materials

Bacterial strains and plasmids

E. coli DH10ß (F⁻ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC)phi80d*lac*ZΔM15 Δ*lac*X74 *deo*R *rec*A1 *end*A1 *ara*Δ139 Δ(*ara, leu*)7697 *gal*U *gal*K lambda⁻ *rps*L *nup*G; Gibco-BRL, Gaithersburg, MD) was used as the cloning host. Plasmids pUC18 (Ap^R ; Vieria and Messing, *Gene* **19** : 259-268, (1982)), pOK12 (Km^R; Vieria and Messing, *Gene* **100** : 189-194 (1991)) and pSL1180 (Ap^R; Brosius, *DNA* **8** : 759, (1989)) were used as cloning vectors. Plasmid pEBCtac (Ap^R Tc^R $lacI^q$ *tac*, shown in Figure 11, was used to overexpress the *Sphingomonas dszB* in *E. coli*.

Media and Reagents

Luria broth (LB) medium was routinely used to propagate *E. coli*. LB medium is 1% tryptone (Difco), 0.5% yeast extract (Difco) and 0.5% NaCl. Rich medium (RM) was used to propagate *Sphingomonas* strain AD109. RM medium is 0.8% nutrient broth, 0.05% yeast extract and 1% glucose. 2YT medium, used in gene expression studies, is 1.6% tryptone, 1% yeast extract and 0.5% NaCl. Basal salts medium (BSM-glucose) contained the following (per liter): phosphate buffer 100 mmol (pH 7.2); glucose, 20 g: NH_4Cl, 2 g; $MgCl_2 \cdot 6H_2O$, 644 mg; $MnCl_2 \cdot 4H_2O$, 1 mg; nitriloacetic acid, 0.1 g; $FeCl_2 \cdot 4H_2O$, 2.6 mg; $Na_2B_4O_7 \cdot 10H_2O$, 0.1 mg; $CuCl_2 \cdot 2H_2O$, 0.15 mg; $Co(NO_3)_2 \cdot 6H_2O$, 0.125 mg; $ZnCl_2$, 2.6 mg; $CaCl_2 \cdot 2H_2O$, 33 mg; $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$, 0.09 mg; and EDTA, 1.25 mg. When required the sulfur source was either 2 mM $MgSO_4$, 300 μM Dibenzothiophene (DBT), 300 μM Dibenzothiophene sulfone

($DBTO_2$) or 300 μM 2-(2-hydroxyphenyl) benzenesulfinate (HPBS). For solid media, agar or agarose was added at a concentration of 1.5% (wt/wt). The antibiotic concentrations for *E. coli* were as follows: ampicillin, 100 μg/ml; kanamycin, 30 μg/ml; tetracycline, 10 μg/ml.

DNA Methods

Restriction enzymes and T4 DNA ligase were purchased from New England Biolabs, Inc. (Beverly, MA) and used as recommended by the supplier. Chromosomal DNA was isolated by the method described by Woo *et al.*, *BioTechniques* **13**: 696-698 (1992). Small scale plasmid preparations from *E. coli* were carried out as described by Birboim and Doly, *Nuc. Acids Res.* **7** : 1513-1523 (1979). Larger scale DNA preparations were carried out with Midi-prep columns from Qiagen (Chatsworth, CA). DNA fragments were purified from agarose gels after electrophoretic separation by the method of Vogelstein and Gillespie (*Proc. Natl. Acad. Sci. USA* **76**: 615-619 (1979). DNA fragments were cloned into vectors by using techniques described by Sambrook *et al.*

Degenerate oligonucleotide probes were end-labeled using standard digoxygenin protocols according to the Boehringer Mannheim DIG Oligonucleotide 3'-End Labeling Kit (Cat. No. 1362372). Hybridization was performed in 5X SSC with blocking solution containing 50% ultrapure deionized formamide at 42°C overnight (16 hr). Detection of hybrids was by enzyme immunoassay according to the Boehringer Mannheim Nonradioactive DIG DNA Labeling and Detection Kit (Cat. No. 1093657).

DNA samples were sequenced by SeqWright (Houston, TX) using a dye-terminator cycling sequencing kit from Perkin Elmer and the 373A and 377 ABI automatic DNA sequencer. The sequence was extended by synthesizing overlapping

oligonucleotides to previously read sequence. The synthesized oligonucleotides were used as primers for continuing sequence reactions. Sequencing reads were assembled and edited to 99.99% accuracy using Genecode's *Sequencher*, version 3.0 computer software.

DNA and protein sequence analysis was performed with the MacVector software program (Oxford Molecular Group, Campbell, CA). Nucleotide and amino acid sequences were compared to sequences in the available databases using BLAST. The Wisconsin Genetics Computer Group (GCG) software (Devereux *et al.*, *Nucl. Acids Res.* **12** : 387-395 (1984)) program GAP was used to generate comparisons of the protein sequences.

Transformation of *E. coli*

Plasmid DNA was introduced into *E. coli* DH10ß by electroporation. Competent ElectroMAX DH10ß (Gibco-BRL, Gaithersburg, MD) were used according to the manufacturer's suggestions.

Preparation of cell-free extracts

Cells grown in the appropriate medium were concentrated to an optical density at 600 nm of 50 by centrifugation and resuspended in 10 mM phosphate buffer (pH 7.0). Cells were disrupted in a French press and debris was removed by centrifugation at 32,000 x g for 20 min. Cell lysates were stored on ice at 4°C.

Desulfurization assays and analytical analysis

HPBS desulfinase activity was assayed by the ability of cell-free lysates to convert HPBS (substrate) to 2-HBP (product) in a one hour assay at 30°C. The amounts of product made and substrate consumed during the reaction

were quantitated by high-pressure liquid chromatography (HPLC) analysis. HPBS desulfinase activity was also measured by fluorescence spectroscopy. In a typical enzyme assay, enzyme activity is determined by the change in fluorescence at an excitation wavelength of 288 nm and an emission wavelength of 414 nm as HPBS is converted to 2-HBP. The assay is initiated by the addition of 20 - 100 μg total protein to a 3 mL solution of 200 μM HPBS in 50 mM phosphate buffer pH 7.5 containing 0.1 M NaCl.

Expression studies

E. coli DH10ß harboring the *Sphingomonas dszB* overexpression plasmid pDA296 was inoculated into 100ml of 2YT medium containing ampicillin and allowed to grow with shaking at 30°C. At an OD_{600} of approximately 0.3, the culture was divided into two parts. One half of the culture was induced by the addition of isopropylthio-β-galactoside (IPTG) (final conc. 1 mM) and the remaining culture was used as an uninduced control (no IPTG was added). Following incubation for an additional 3 hr, both cultures were harvested and cell-free lysates were prepared.

Protein purification and N-terminal sequencing

Sphingomonas AD109 cell paste was resuspended in an approximately equal weight of 25 mM phosphate buffer pH 7.5 containing 0.1 mM EDTA, 0.5 mM dithiothreitol (DTT), 10 μg/mL DNAse and 1 mM phenylmethylsulfonyl fluoride and passed through a French press mini-cell at about 20,000 psi. Cell debris was removed by centrifugation and the cell lysate was fractionated over an Econo-Pac High Q cartridge manufactured by Bio-Rad. A linear 0-0.5 M NaCl gradient was used to elute the bound protein into

fractions. The active fractions were identified by a 2-HBP fluorescence enzyme assay (excitation/emission wavelengths set at 288/414 nm). The active fractions were pooled and desalted over a Bio-Rad P6 gel filtration cartridge, diluted to 1.7 M ammonium sulfate and fractionated over a Phenyl Superose HR 5/5 column manufactured by Pharmacia. A linear 1.7-0.0 M ammonium sulfate gradient was used to elute protein into fractions. Active fractions were identified and pooled as described above. Identity and purity of the AD109 HPBS desulfinase protein was also determined by SDS-PAGE and Western blots using antibodies generated against the DszB protein from *Rhodococcus erythroplis* strain IGTS8. N-terminal microsequencing of the HPBS desulfinase was carried out by Edman degradation after transfer of the purified protein to a polyvinylidene difluoride (PVDF) membrane.

SDS-PAGE and Western Blot Analysis

Protein separations were done with Novex (San Diego, CA) precast 10% polyacrylamide gels with Tris-Glycine-sodium dodecyl sulfate (SDS) (Laemmli) running buffer. Western blot analysis was carried out by first transferring the proteins electrophoretically to nitrocellulose membranes as recommended by Biorad (Hercules, CA). Blots were treated with antisera raised against the purified IGTS8 DszB protein (primary antibody) and then with goat anti-rabbit antisera conjugated to horseradish peroxidase as the secondary antibody. Finally, the proteins were detected with a horseradish peroxidase catalyzed chemiluminescent reaction.

Example 1 Soil enrichments and isolation of a microorganism that can use HPBS as a sole sulfur source

Three independent soil samples from oil-contaminated sites were used to perform soil enrichments for microorganisms able to use HPBS as a sole sulfur source. Approximately 5 grams of each soil sample was placed into a sterile 250 ml flask along with 50 ml of BSM Glucose medium containing HPBS (300 μM) as the sole source of sulfur. Following incubation for 96 hrs at 30°C, a 3 ml sample of each enrichment was transferred to fresh BSM Glucose medium containing HPBS. After 72 hrs, one of the three flasks (flask #3) showed visible turbidity, while the two remaining flasks showed no visible increase in turbidity (even after more than a week of incubation). Microscopic analysis of the contents of flask #3 revealed the presence of a mixed population of bacterial cells (i.e., sessile and motile rods of varying shapes; large and small coccoid shaped bacteria). After repeated liquid subculture enrichments with HPBS as the sole sulfur source, the contents of the flask was plated onto several RM and LB agar plates. Following incubation at 30°C, a variety of microorganisms with different colony morphologies was present. Analysis of individual colonies from these plates identified a pure isolate that efficiently used HPBS as a sole sulfur source. This strain, designated AD109, was selected for further analysis.

Example 2 Characterization and identification of strain AD109

The HPBS utilizing strain AD109 is a Gram-negative, motile rod that forms distinctive yellow colonies on agar plates. It grows somewhat poorly on LB agar, but grows

rather well on RM agar plates. Like *Rhodococcus* IGTS8, strain AD109 also has the ability to produce clearing zones on a BSM Glucose DBT-sulfone plate. The optimal growth temperature of AD109 was found to be between 30 and 37°C.

Based on fatty acid analysis (Acculab, Inc., Newark, DE), this strain was identified as a *Sphingomonas* species. Strain AD109 was a "good" match to *S. paucimobilis* (formerly *Pseudomonas paucimobilis*) based on its "similarity index". The similarity index is a mathematical expression of the extent to which the fatty acid profile of a given unknown matches the mean profile for an organism in the TSBA database. Strain AD109 had an index value of 0.426 which indicates that it is from a strain of a species that differs significantly from those represented in the database. A similarity index of 0.5 or above is considered to be an "excellent" match (a value of 1.0 being the highest possible). On the other hand, an index below 0.3 indicates that the sample is from a species that is not likely to be in the database. Based on 16S rRNA sequence analysis and the presence of sphingoglycolipids, Yabuuchi *et al.* (*Microbiol. Immunol.* **34** : 99-119 (1990)) proposed that *P. paucimobilis* be reclassified and placed into the genus *Sphingomonas*.

Example 3 Growth characteristics of *Sphingomonas* species strain AD109

Evidence for the existence of an HPBS desulfinase activity was demonstrated by monitoring the supernatant of a AD109 culture growing in BSM Glucose HPBS (300 μM). By the time the culture was well into stationary phase all of the HPBS had been converted with no apparent accumulation of identifiable intermediates. There was, however, a

transient production of a small amount of 2-HBP, as determined by HPLC analysis, which also disappeared with time. This preliminary result suggested that AD109 may also be capable of metabolizing 2-HBP. *Sphingomonas* strain AD109 was also capable of utilizing DBT-sulfone ($DBTO_2$) as a sole sulfur source. The ability to utilize DBT-sulfone as a sole sulfur source suggests that strain AD109 may also contain a gene that encodes DBT-sulfone monooxygenase activity.

During the course of growth studies it was discovered that strain AD109 could utilize DBT as a sole sulfur source. While growing with DBT, however, the culture supernatant takes on a very characteristic orange/brown color with an absorption maximum of approximately 470 nm. Orange-colored oxidation products have been previously identified in a number of *Pseudomonas* species that are capable of degrading DBT (Monticello *et al.*, *Appl. Environ. Microbiol.* **49** : 756-760 (1985)); Foght and Westlake, *Can. J. Microbiol.* **36** : 718-724 (1990)). No such color development was detected in cultures growing with either HPBS or DBT-sulfone as sulfur sources.

Example 4 Demonstration of HPBS desulfinase activity in AD109 cell-free lysates

A cell-free lysate prepared from a culture of *Sphingomonas* strain AD109 (grown in BSM Glucose medium containing HPBS) was used in a time course study to examine the rate at which HPBS is converted to 2-HBP. As presented in Figure 4, at a protein concentration of 4 mg/ml there was a linear increase in 2-HBP production and a concomitant disappearance of HPBS.

The product of the *in vitro* reaction was confirmed to be 2-HBP by a spectral comparison to authentic 2-HBP. The ultraviolet absorption spectrum of the suspected 2-HBP peak produced by the action of the AD109 lysate is virtually identical with that of the 2-HBP standard. Furthermore, the molecular weight of the unknown compound was exactly that of authentic 2-HBP as determined by GC-MS analysis.

Example 5 Purification of the HPBS desulfinase from *Sphingomonas* AD109

HPBS desulfinase was purified from AD109 by a series of chromatographic steps using a Bio-Rad low pressure column chromatography Econo system and a Pharmacia FPLC (Gray *et al.*, *Nature Biotech.* **14** : 1705-1709 (1996)). The steps included fractionation over an anion exchange resin followed by a hydrophobic interaction column chromatography step. These protein purification steps are described above. A 15-20 fold purification was achieved in these two steps which is comparable to protein preparations from a *Rhodococcus* IGTS8 lysate.

The molecular weight of this protein by SDS-PAGE was estimated to be 40,000 daltons, which is approximately the same size as DszB purified from IGTS8. Western analysis demonstrated that the purified protein shows some cross-reactivity with anti-DszB antisera.

Nonlinear regression analysis of an enzyme progress curve was performed according to the general method described by Duggleby, *Methods Enzymol.* **249** : 61-90 (1995). The analysis involves fitting the integrated Michealis-Menton rate equation $V_m^*t = y-K_m^*\ln(1-y/[A]_0)$ to concentration vs. time data from the enzyme catalyzed reaction of 2-(2-phenyl)benzenesulfinate to 2-

hydroxybiphenyl monitored to completion by fluorescence. The semi-pure protein sample was generated by fractionation of a crude lysate over Q Sepharose Fast Flow resin (Pharmacia) by a linear 0-0.5 M NaCl gradient, as discussed in more detail above. The purity of the active fraction was determined by SDS-PAGE. Pure enzyme is not necessary for the application of enzyme progress curve analysis, however, the calculation of k_{cat} ($V_m = [E]_t * k_{cat}$) was limited to a value range as only a crude estimate of the enzyme concentration was available. The reaction conditions were as follows. A 3 mL reaction solution containing 1 μM HPBS and 0.1 M NaCl in 50 mM phosphate at pH 7.5 and 30°C was initiated by the addition of 0.023 mg total protein and was monitored for 30 min by fluorescence at an excitation wavelength of 288 nm and an emission wavelength of 414 nm. The data were fit to the equation using the Kaleidagraph data analysis/graphics application (Abelleck Software).

Based on the kinetic parameters calculated from the enzyme progress assay (K_m=0.3 μM and V_m=0.1 μM/min), the minimum k_{cat} = 0.5 min^{-1}. However, a more realistic value would be on the order of 2 min^{-1} in view of the fact that the preparation is estimated to be about 25% pure. Therefore, the HPBS desulfinase from *Sphingomonas* AD109 appears to be comparable to that from *Rhodococcus* IGTS8 with the possibility of a higher catalytic efficiency (k_{cat}/K_m).

The N-terminal amino acid sequence of the purified *Sphingomonas* HPBS desulfinase was also determined. Protein microsequencing using standard methods of analysis resulted in the following amino acid sequence:

1 10 20

TTDIHPASAA SSPAARATIT YS (SEQ ID NO.: 7)

A comparison of the putative AD109 HPBS desulfinase N-terminal sequence with that of the N-terminus of the IGTS8 DszB protein revealed that 9 out the 22 amino acid residues were identical (41%). In order to determine whether the purified protein is, in fact, the *Sphingomonas* desulfinase protein, a degenerate (192 permutations) 17-mer oligonucleotide probe with the following sequence: 5' ACN GAY ATH CAY CCN GC 3' (SEQ ID NO.: 8), was designed based on the determined N-terminal sequence. Following labeling with a non-isotopic label this probe was used in hybridization studies using the cloned *Sphingomonas* AD109 HPBS desulfinase gene (see below) and the *dszB* gene from IGTS8 (Denome *et al.*, *J. Bacteriol.* **176** : 6707-6716 (1994); Piddington *et al.*, *App. Environ. Microbiol.* **61** : 468-475 (1995). The labeled oligonucleotide probe hybridized to the cloned *Sphingomonas* HPBS desulfinase gene which indicated that the correct protein had been purified. However, no signal was detected in the lane containing a fragment harboring the *Rhodococcus dsz* B gene.

Example 6 Cloning of the *Sphingomonas* AD109 HPBS desulfinase gene

Strain AD109 has been shown to be capable of using HPBS as a sole sulfur source and clearing a $DBTO_2$ plate. On the assumption that the gene(s) responsible for $DBTO_2$ clearing and HPBS desulfinase activity are genetically closely linked, as they are in *Rhodococcus* IGTS8, a cloning scheme was devised to isolate the HPBS desulfinase gene from *Sphingomonas* strain AD109. Total genomic DNA from strain AD109 was digested with either *Eco*RI, *Bam*HI, and *Hind*III and the resulting fragments were ligated into pUC18

or pSL1180. Following transformation of *E. coli* DH10ß, approximately 1000-2000 Lac-negative, ampicillin-resistant colonies of each library were screened for the ability to clear a $DBTO_2$ plate. No clearing colonies were detected amongst transformants derived from either the *Eco*RI or *Bam*HI libraries. However, two clearing colonies were detected utilizing the *Hin*dIII library and one clearing colony was detected with the *Not*I library. Based on restriction endonuclease profiles, both colonies from the *Hin*dIII library contained the same large fragment (~20 kb). Furthermore, there was measurable HPBS desulfinase activity in cell-free lysates of these strains.

The single clearing colony from the *Not*I library contained a 6.5 kb fragment which, according to restriction endonuclease mapping, overlapped the 20 kb *Hin*dIII fragment. This clone also contained measurable HPBS desulfinase activity.

Subcloning analysis localized the genes responsible for $DBTO_2$ clearing and HPBS desulfinase activity to a 6 kb *Hin*dIII-*Not*I fragment. A smaller 2.7 kb *Hin*dIII-*Sma*I fragment was subsequently found to retain HPBS desulfinase activity, but lost the ability to clear a $DBTO_2$ plate. It is likely, therefore, that the gene that confers the ability to produce clearing zones on a DBT-sulfone plate spans the *Sma*I site.

Example 7 DNA sequence analysis of the *Sphingomonas* sp. strain AD109 desulfurization gene cluster

The nucleotide sequence of a 4144 bp region which encompasses the AD109 HPBS desulfinase gene was

determined from both DNA strands and is present in Figure 6 (SEQ ID NO.: 12). The overall G+C content of the first 3837 base pairs of the AD109 sequence is 64.5%, a value which is consistent with the range of G+C values (61.7 - 67.2%) reported for various *Sphingomonas* species (Yabuuchi *et al.* (1990)). A comparison of the AD109 nucleotide sequence with the IGTS8 *dsz* sequence by DNA matrix analysis revealed that a considerable amount of homology exists between the two sequences as evidenced by the presence of a near continuous diagonal line.

Open reading frame analysis of the AD109 sequence revealed the presence of a number of ORFs on both DNA strands, but of these, only three contained the codon-choice pattern characteristic of microorganisms with G-C rich genomes (West *et al.*, *Nucl. Acids Res.* **16:** 9323-9334 (1988)). All three identified ORFs were in the same transcriptional orientation. A strong preference for codons with either G or C occurred in positions 1 and 3. The first codon position of all three ORFs ranged from 67 to 72%, while the third codon position of all three ORFs ranged from 79-81%. In addition, the predicted translation initiation sites of all three ORFs are preceded by sequences that resemble a consensus ribosome binding site.

The entire nucleotide sequence of the AD109 region was used to conduct a BLAST search of the available DNA databases. The *Rhodococcus* IGTS8 *dsz* genes were the highest scoring sequences that demonstrated homology to the *Sphingomonas* sequences. The only other nucleotide sequence that demonstrated any significant homology to the *Sphingomonas* DNA, was the *Streptomyces pristinaespiralis snaA* gene which encodes the large

subunit of the PII_A synthase (Blanc *et al.*, *J. Bacteriol.* **177** : 5206-5214 (1995)). The *Sphingomonas dszA* and *S. pristinaespiralis snaA* genes demonstrate about 60% identity over a 800 bp region proximal to the 5' end of each gene.

The first ORF (bp 442-1800; Figures 1A-1D) is 71% identical (at the nucleotide level) to the *Rhodococcus dszA* gene. The primary translation product of ORF-1 would encode a protein (*Sphingomonas* DszA or Dsz(S)) that contains 453 amino acids with a predicted molecular weight of 50,200. More importantly, this protein demonstrates considerable homology to the amino acid sequence of *Rhodococcus* DszA (Dsz(R), SEQ ID NO.: 9) over the entire length of the polypeptide (76% identity and 87% similarity; Figure 8). The protein databases were also searched with the *Sphingomonas* DszA protein sequences. Aside from the DszA protein of *Rhodococcus* IGTS8, several other proteins demonstrated significant homology to the *Sphingomonas* DszA protein. These include a hypothetical 49.3 kD protein in the IDH-DEOR intergenic region of *Bacillus subtilis* which showed 45% identity over 382 residues, the PII_A synthase SnaA subunit of *S. pristineaspiralis* (Blanc *et al.*, *J. Bacteriol.* **177** : 5206-5214 (1995)) which was 49% identical over 358 residues and the nitrilotriacetate monooxygenase of *Chelatobacter heintzii* (Xu *et al.*, Abstracts of the 95th General Meeting of the American Society for Microbiology, Q-281) which was 50% identical over the 335 residues examined.

The stop site of the *Sphingomonas* ORF-1 shows a 4-bp overlap with the translation start site of the second ORF (bp 1800-2906; Figures 2A-2C), which shows a

high degree of homology to the *Rhodococcus* IGTS8 *dszB* gene (67% identity). It was determined that the primary translation product of ORF-2 would encode a 369-amino acid polypeptide with a predicted molecular weight of 40,000 (*Sphingomonas* DszB or Dsz(S)). At the amino acid level this putative protein is 66% identical (75% similarity) to the *Rhodococcus* HPBS desulfinase protein DszB (DszB(R), SEQ ID NO: 10), as shown in Figure 9. Except for the IGTS8 DszB protein, a BLAST search with the *Sphingomonas* DszB sequence did not identify any other significant homologous sequences in the available databases. The predicted N-terminus of the *Sphingomonas* DszB protein matches identically the N-terminus of the HPBS desulfinase purified from AD109 cell lysates, except that the amino-terminal methionine was absent. Removal of the methionine residue has been shown to occur when the second amino acid is Ala, Ser, Gly, Pro, Thr or Val (Hirel *et al.*, *Proc. Nat. Acad. Sci. USA* **86** : 8247-8251 (1989)).

The stop site of the *Sphingomonas dszB* gene also shows a 4-bp overlap with the translation start site of the third ORF. This ORF (bp 2906-4141; Figures 3A-3C), shows significant homology to the *Rhodococcus* IGTS8 *dszC* gene. For example, over the first 931 bp, this ORF is 69% identical to the IGTS8 *dszC* gene and the N-terminus of the protein predicted by this sequence (*Sphingomonas* DszC, DszC(S)) is 67% identical to the N-terminus of *Rhodococcus* DszC (DszC(R), SEQ ID NO: 11), as shown in Figure 10. A BLAST search of the protein databases with the available *Sphingomonas* DszC sequence identified a number of proteins in addition to the IGTS8 DszC protein. The *Sphingomonas* DszC protein is 32% identical

(over 199 residues) to Isobutylamine *N*-Hydroxylase (IBAH) of *Streptomyces viridifaciens*. It has previously been shown that IBAH exhibits the greatest similarity to the IGTS8 DszC protein (Parry *et al.*, *J. Bacteriol.*, **179**: 409-416 (1997)). In addition, the AD109 DszC protein showed variable homology to a number of acyl coenzyme A dehydrogenases. For example, the N-terminal 300 residues of the *Sphingomonas* DszC protein is 29% identical to the acyl CoA dehydrogenase of *B. subtilis*.

The sequences (400 bp) directly upstream of the *dszA* start site contain regulatory elements (i.e., promoter elements) that control transcription of the AD109 *dsz* gene cluster. A comparison of this potential promoter region with the IGTS8 *dsz* promoter region failed to reveal any significant homology. It has been shown that the IGTS8 *dsz* promoter region encompasses a region of potential diad symmetry that may contain an operator (Li *et al.*, *J. Bacteriol.* **178** : 6409-6418 (1996)). An examination of the AD109 sequences directly upstream of *dszA* revealed no such palindromic sequence.

Example 8 Expression of the *Sphingomonas dszB* gene in *E. coli*

The AD109 *dszB* gene was subcloned into the *tac* promoter expression vector, pEBCtac, in two steps. The first step involved cloning a 1.2 kb *Pst*I-*Bgl*II fragment that contained the entire coding region of the AD109 *dszB* gene (Figures 2A-2C) into the polylinker plasmid pOK12. The resulting plasmid, designated pDA295, contained a unique *Xba*I site upstream of the *dszB* gene.

In the second step, a 1.2 kb *Xba*I-*Bgl*II fragment from pDA295 that contained the entire *dszB* gene was cloned into the *Xba*I and *Bgl*II sites of pEBCtac, thus placing the AD109 *dszB* gene under the transcriptional control of the *tac* promoter. This plasmid, designated pDA296 and presented in Figure 7, was introduced into *E. coli* DH10ß for expression studies.

HPBS desulfinase assays (2 mg/ml protein) using cell-free lysates prepared from induced and uninduced cultures of DH10ß/pDA296 were performed. In the absence of IPTG the cell-free lysate contained very little HPBS desulfinase activity. Only 22 nmoles of 2-HBP were produced during the 60 min. incubation period which corresponds to a specific activity of 0.2 (nmoles 2-HBP formed/min/mg protein). The lysate prepared from the IPTG-induced culture, however, had approximately 20 times more HPBS desulfinase activity (4.2 nmoles 2-HBP formed/min/mg protein) than the lysate prepared from the uninduced culture.

EQUIVALENTS

Those skilled in the art will know, or be able to ascertain, using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. These and all other equivalents are intended to be encompassed by the following claims.